UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/674,573	09/30/2003	Liam D. Comerford	YOR920030388US1 (163-12)	3846

24336 7590 09/04/2008
KEUSEY, TUTUNJIAN & BITETTO, P.C.
20 CROSSWAYS PARK NORTH
SUITE 210
WOODBURY, NY 11797

EXAMINER	
LENNOX, NATALIE	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
09/04/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	**Application No.** 10/674,573	**Applicant(s)** COMERFORD, LIAM D.	
	Examiner NATALIE LENNOX	**Art Unit** 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _05 March 2008_.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-36_ is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-36_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

This Office Action has been issued in response to the Pre-Appeal Brief Request for Review filed on March 5, 2008.

Response to Arguments

1. Applicant's arguments, see Pre-Appeal Brief Request for Review pages 1-3, filed March 5, 2008, with respect to the rejection(s) of claim(s) 1 and 16 under §102 (e) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Walker et al. (US Patent 6,434,529). Walker's Col. 5, lines 49-60 and the example provided in Col. 6, lines 34-48, wherein from Col. 5, lines 56-60 it is stated that the recognition result contains the tokens or words the user said (decoding at least one word in acoustic data representing an acoustic signal that comprises a human utterance and determining acoustic word boundaries within the acoustic data). From the example provided in Col. 6, lines 36-48, Walker extracts the command of "I want a (hamburger|burger) with <toppings>" or rule <order>, which is determined by previously generating the recognition result identifying the words the user said, and where it is clearly presented that the rules or object instances <toppings>, <condiment>, and <veggy> relate to the acoustic data segments identified from the user utterance. Other examples are provided in Col. 4, lines 34-40, wherein the commands are <play>, <stop>, and <goto> and <lineno> represent the acoustic data segment from the utterance.

2. Applicant's arguments, see Pre-Appeal Brief Request for Review page 4, filed March 5, 2008, with respect to the rejection(s) of claim(s) 32 under 35 U.S.C. §103 (a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Stammler et al. (US Patent 6,839,670). Stammler's Col. 9, lines 43-51 and Col. 5, lines 36-41 clearly provide examples where a command is processed with a speaker-independent vocabulary and accordingly the audio data is processed by a speaker-dependent vocabulary.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claims *** are rejected under 35 U.S.C. 102(b) as being anticipated by Walker et al. (US Patent 6,434,529), hereinafter Walker.

As per claims 1 and 15, Walker teaches a method and program storage device readable by machine (Col. 19, lines 25-37), for extracting commands and acoustic data in a same utterance, comprising the steps of:

decoding at least one word in acoustic data representing an acoustic signal that comprises a human utterance and determining acoustic word boundaries within the

acoustic data (Col. 5, lines 49-60 and the example provided in Col. 6, lines 34-48, wherein from Col. 5, lines 56-60 it is stated that the recognition result contains the tokens or words the user said);

extracting at least one command in a decoded utterance (Col. 6, lines 36-48, Walker extracts the command of "I want a (hamburger|burger) with <toppings>" or rule <order>, which is determined by previously generating the recognition result identifying the words the user said); and

identifying acoustic data segments in the utterance based on the acoustic word boundaries (Col. 6, lines 36-48, Walker clearly presents that the rules or object instances <toppings>, <condiment>, and <veggy> relate to the acoustic data segments identified from the user utterance. Other examples are provided in Col. 4, lines 34-40, wherein the commands are <play>, <stop>, and <goto>, and the label <lineno> represents the acoustic data segment from the utterance.

As per claim 3, Walker teaches the method as recited in claim 1, further comprising the step of executing the at least one command from the decoded utterance (Col. 1, lines 17-20, and Col. 4, lines 55-58).

As per claim 5, Walker teaches the method as recited in claim 3, further comprising the step of submitting at least one non-command voice data segment for recognition using the recognizer vocabulary (Col. 5, lines 49-60, and Col. 6, lines 36-48, wherein Walker clearly presents that the rules or object instances <toppings>,

<condiment>, and <veggy> relate to the acoustic data segments identified from the user utterance. Other examples are provided in Col. 4, lines 34-40, wherein the commands are <play>, <stop>, and <goto> and <lineno> represent the acoustic data segment from the utterance.).

As per claim 7, Walker teaches the method as recited in claim 1, further comprising the step of submitting the acoustic data segments for recognition when computing resources are available (Col. 5, lines 49-60, and Col. 6, lines 36-48, wherein Walker clearly presents that the rules or object instances <toppings>, <condiment>, and <veggy> relate to the acoustic data segments identified from the user utterance. Other examples are provided in Col. 4, lines 34-40, wherein the commands are <play>, <stop>, and <goto>, and the label <lineno> represents the acoustic data segment from the utterance. Also, Col. 12, lines 14-29).

As per claim 8, Walker teaches the method as recited in claim 1, wherein the step of extracting at least one command from the utterance includes employing one or more grammars to distinguish the command (Col. 12, lines 14-29 and Figure 1, elements 12 (12a, 12b, and 12c).

As per claims 9 and 27, Walker teaches the method as recited in claims 8 and 25, wherein the grammars include a form for extracting information for an order or verbal contract (Walker et al. teach a system (Fig. 1) that includes result listener 18,

parse tree 20, and a tags parser 24. The result listener receives the recognition result and uses the grammar from grammars 12, which includes the rule that was matched to turn the result into a parse tree 20 (Col. 5, lines 61-63), then the tags parser 24 evaluates the parse tree 20 and creates an object instance, called a rule object, for each rule it encounters in the parse tree 20. The name of a rule object for any given rule is, for purposes of example, of the form $name. That is, the name of the rule object is formed by prepending a '$' to the name of the rule (Col. 6, lines 14-19). In a specific example, Col. 6, lines 36-44 describe an example of a form (or rule) for a food order).

As per claim 12, Walker teaches the method as recited in claim 8, wherein the step of using grammars includes the step of associating at least one grammar label with the corresponding segment of acoustic data that has been decoded into a command (Col. 6, lines 36- 44, give an example of a user's utterance "I want a burger with onions and mustard," wherein the label "<veggy>" is associated with the recognized acoustic data "onions" and label "<order>" with "I want a (hamburger|burger) with <toppings>," etc.).

As per claim 16 and 31, Walker teaches a method and program storage device readable by machine (Col. 19, lines 25-37), for recognizing at least one command and at least one segment of acoustic voice data in a same utterance comprising the steps of:

decoding at least one word in voice data representing the acoustic signal that

comprises a human utterance and determining the acoustic word boundaries within the voice data (Col. 5, lines 49-60 and the example provided in Col. 6, lines 34-48, wherein from Col. 5, lines 56-60 it is stated that the recognition result contains the tokens or words the user said);

extracting at least one command from the utterance (Col. 6, lines 36-48, Walker extracts the command of "I want a (hamburger|burger) with <toppings>" or rule <order>, which is determined by previously generating the recognition result identifying the words the user said); and

associating segments in the voice data based on the acoustic word boundaries with labels (Col. 6, lines 36-48, Walker clearly presents that the rules or object instances <toppings>, <condiment>, and <veggy> relate to the acoustic data segments identified from the user utterance. Other examples are provided in Col. 4, lines 34-40, wherein the commands are <play>, <stop>, and <goto>, and the label <lineno> represents the acoustic data segment from the utterance.

As per claim 17, Walker teaches the method as recited in claim 16, wherein the step of extracting includes employing an application, which identifies commands in the utterance in accordance with the labels (Col. 4, lines 29-31 and Col. 4, lines 34-45). The application program may be referenced directly from scripting language within the tags (labels) defined by the rule grammar (Col. 4, lines 29-31). A portion of the rule grammar for the example of the media player is shown on Col. 4, lines 34-40, where commands such as "play," "go," and "start" are labeled <play>. Also the label <play> is part of the

rule grammar for <command>. A tags parser program is invoked to interpret the tags in a recognition result matching one of the rules, such as Processing of recognition results in the application programs may be simplified to an invocation of the tags parser (Col. 4, lines 41-45).

As per claim 24, Walker teaches the method as recited in claim 16, further comprising the step of buffering the utterance to be processed and maintaining the utterance in memory during processing of the utterance (Fig. 8 and Col. 14, lines 57-58 and 62-64). "SUSPENDED" state 136 of the Recognizer (Fig. 8), wherein the Recognizer remains in the SUSPENDED state 136 until processing of the result finalization event is completed (Col. 14, lines 57-58). In the SUSPENDED state 136 the Recognizer buffers incoming audio. This buffering allows a user to continue speaking without speech data being lost (Col. 14, lines 62-64).

As per claim 25, Walker teaches the method as recited in claim 16, wherein the step of associating segments includes employing grammars to associate a unique label with each command segment in the utterance (Col. 6, lines 36- 44. The association of the label to the command segment "I want a (hamburger|burger) with "from the user utterance "I want a (hamburger|burger) with onions and mustard." The labels and are also associated with the words onion and mustard, respectively.).

Claim Rejections - 35 USC § 103

5. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

6. Claims 10-11, 13, 26, and 28-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Walker (US Patent 6,434,529).

As per claims 10 and 28, Walker teaches the method as recited in claims 8 and 25, wherein the grammars include a form for reminding a user to perform a task (Walker et al. teach a system (Fig. 1) that includes result listener 18, parse tree 20, and a tags parser 24. The result listener receives the recognition result and uses the grammar from grammars 12, which includes the rule that was matched to turn the result into a parse tree 20 (Col. 5, lines 61-63), then the tags parser 24 evaluates the parse tree 20 and creates an object instance, called a rule object, for each rule it encounters in the parse tree 20. The name of a rule object for any given rule is, for purposes of example, of the form $name. That is, the name of the rule object is formed by prepending a '$' to the name of the rule (Col. 6, lines 14-19). In a specific example, Col. 6, lines 36-44, describe an example of a form (or rule) for a food order. It would have been obvious to one having ordinary skill in the art that this form or rule could also be applied to remind a user to perform a task).

As per claims 11 and 29, Walker teaches the method as recited in claims 8 and 25, wherein the grammars include a form for extracting maximum meaningful length segments under interruption or silence conditions (Walker et al. teach a system (Fig. 1)

that includes result listener 18, parse tree 20, and a tags parser 24. The result listener receives the recognition result and uses the grammar from grammars 12, which includes the rule that was matched to turn the result into a parse tree 20 (Col. 5, lines 61-63), then the tags parser 24 evaluates the parse tree 20 and creates an object instance, called a rule object, for each rule it encounters in the parse tree 20. The name of a rule object for any given rule is, for purposes of example, of the form $name. That is, the name of the rule object is formed by prepending a '$' to the name of the rule (Col. 6, lines 14-19). In a specific example, Col. 6, lines 36-44, describe an example of a form (or rule) for a food order. It would have been obvious to one having ordinary skill in the art that this form or rule could also be applied to extract maximum meaningful length segments under interruption or silence conditions).

As per claim 13, Walker teaches the method as recited in claim 12, wherein the label includes a numerical value associated with each command (Col. 6, lines 36-44, give an example of a user's utterance "I want a burger with onions and mustard," wherein the label "<order>" is associated with the acoustic data segment "I want a (hamburger|burger) with <toppings>." It would have been obvious to a person having ordinary skill in the art to include a numerical value to the label. For example, if there was a rule for another "order" such as "I want a <flavor> ice cream" the label could have included a number "<order2>").

As per claim 26, Walker teaches the method as recited in claim 25, wherein the label includes a numerical value (Col. 6, lines 36-44, give an example of a user's utterance "I want a burger with onions and mustard," wherein the label "<order>" is associated with the acoustic data segment "I want a (hamburger|burger) with <toppings>.");

It would have been obvious to a person having ordinary skill in the art to include a numerical value to the label. For example, if there was a rule for another "order" such as "I want a <flavor> ice cream" the label could have included a number "<order2>").

7. Claims 2, 4, 6, 14, 18-20, 23, and 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Walker (US Patent 6,434,529) in view of Stammler et al. (US Patent 6,839,670), hereinafter Stammler.

As per claims 2 and 30, Walker teaches the method as recited in claims 1 and 16, but does not specifically mention wherein the step of determining acoustic word boundaries includes finding segment boundaries by iteratively comparing the same utterance to a plurality of vocabularies.

However, Stammler does teach wherein the step of determining acoustic word boundaries includes finding segment boundaries by iteratively comparing the same utterance to a plurality of vocabularies (Col. 5, lines 38-41, Col. 2, lines 47-49, Col. 4, lines 60-63, Col. 5, lines 11-13, and Col. 2, lines 61-65, wherein the step of determining acoustic word boundaries includes finding segment boundaries in the speaker independent and speaker dependent vocabularies. The speaker independent

recognizer recognizes general control commands, numbers, names, letters, etc., without requiring that the speaker or user train one or several of the words ahead of time (Col. 4, lines 60-63) and the speaker dependent recognizer recognizes user-specific/speaker-specific names or functions, which the user/speaker defines and trains (C01.5, lines 11-13). The system permits a speech command input or speech dialog control that is for the most part adapted to the natural way of speaking, and an extensive vocabulary of admissible commands that is made available to the speaker for this (Col. 2, lines 61-65). In a specific example (Col. 5, lines 38-41), "call uncle Willi," the speaker independent recognizer recognizes "call" and the speaker dependent recognizer, "uncle Willi.").

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of finding segment boundaries by iteratively comparing the same utterance to a plurality of vocabularies as taught by Stammler for Walker's method because Stammler provides a system that permits a speech command input or speech dialog control that is for the most part adapted to the natural way of speaking, and an extensive vocabulary of admissible commands that is made available to the speaker for this (Col. 2, lines 60-65).

As per claim 4, Walker teaches the method as recited in claim 3, but does not specifically mention further comprising at least one of storing the acoustic data segments and using the acoustic data segments in executing the at least one command.

However, Stammler does teach further comprising at least one of storing the acoustic data segments and using the acoustic data segments in executing the at least one command (Col. 5, lines 36-41, Col. 4, lines 55-57, and Col. 5, lines 11-18. The step of storing the acoustic data segments is done by the speaker-dependent recognizer, which "the user/speaker defines and trains" with "user-specific/speaker-specific names or functions" (the names or functions are the acoustic data segments added to the speaker dependent vocabulary) (Col. 5, lines 11-18). The step of using the acoustic data segments in executing the at least one command is demonstrated as an example when the user utters the command "call uncle Willi." The speaker-independent vocabulary recognizes the command "call" and the speaker-dependent vocabulary the acoustic data segment "uncle Willi" (Col. 5, lines 36-41). Clearly the command "call" needs the acoustic data segment "uncle Willi" in order to execute the complete command).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of storing data segments and using the data segments in executing the at least one command as taught by Stammler for Walker's method because Stammler provides the speaker dependent recognizer so that the user/speaker has the option of setting up or editing personal vocabulary and adapting this vocabulary at any time to accommodate his/her needs (Col. 5, lines 13-18).

As per claim 6, Walker teaches the method as recited in claim 1, but does not specifically mention further comprising the step of changing a recognizer vocabulary.

However, Stammler does teach further comprising the step of changing a recognizer vocabulary (Col. 5, lines 37-41). In a specific example, in order to recognize the complete command "call uncle Willi," the word "call" would be recognized by the speaker-independent vocabulary and "uncle Willi" would be recognized by the speaker-dependent vocabulary).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of changing a recognizer vocabulary as taught by Stammler for Walker's method because Stammler speaker dependent vocabulary has the option for a user setting up or editing a personal vocabulary with data that fits his/her needs (Col. 5, lines 13- 18) and the speaker independent vocabulary only contains general control commands, numbers, names, letters, etc., already trained and without being able to be modified by the user (Col. 4, lines 60-63, and Col. 5, lines 8-10).

As per claim 14, Walker teaches the method as recited in claim 1, but does not specifically mention further comprising the step of executing the at least command in the utterance using undecoded acoustic data from within the same utterance (Col. 4, lines 60-62 and Col. 9, lines 19-29). Speaker independent recognizer is capable of recognizing general control commands, numbers, names, letters, etc. (Col. 4, lines 60-62) from an utterance even when the utterance contains garbage words ("non-words")

or unnecessary information. (Col. 9, lines 19-29, for example command: "circle with radius one" from utterance: "I now would like to have a circle with radius one," wherein "I now would like to have a..." is interpreted as undecoded acoustic data.).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of executing the at least command in the utterance using undecoded acoustic data as taught by Stammler for Walker's method because Stammler provides a classification unit for the speaker independent recognizer (Fig. 2) that is able to recognize and separate filler phonemes or garbage words. Garbage words are language complements, which are added by the speaker - unnecessarily - to the actual speech commands, but which are not part of the vocabularies of the speech recognizer (Col. 9, lines 18-25)).

As per claim 18, Walker teaches the method as recited in claim 16, but does not specifically mention further comprising the step of executing the at least one command utilizing undecoded information in the acoustic voice data.

However, Stammler does teach further comprising the step of executing the at least one command utilizing undecoded information in the acoustic voice data (Col. 4, lines 60-62 and Col. 9, lines 19-29). Speaker independent recognizer is capable of recognizing general control commands, numbers, names, letters, etc. (Col. 4, lines 60-62) from an utterance even when the utterance contains garbage words ("non-words") or unnecessary information. (Col. 9, lines 19-29, for example command: "circle with

radius one" from utterance: "I now would like to have a circle with radius one," wherein "I now would like to have a..." is interpreted as undecoded information.).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of executing the at least command in the utterance using undecoded acoustic data as taught by Stammler for Walker's method because Stammler provides a classification unit for the speaker independent recognizer (Fig. 2) that is able to recognize and separate filler phonemes or garbage words. Garbage words are language complements, which are added by the speaker – unnecessarily- to the actual speech commands, but which are not part of the vocabularies of the speech recognizer (Col. 9, lines 18-25).

As per claim 19, Walker teaches the method as recited in claim 16, but does not specifically mention wherein the step of extracting includes the step of storing at least one non-command voice data segment.

However, Stammler does teach wherein the step of extracting includes the step of storing at least one non-command voice data segment (Col. 5, lines 11-15 and Col. 5, lines 36-41). The speaker-dependent recognizer is capable of storing "user-specific/speaker- specific names or functions, which the user/speaker defines and trains. The user/speaker has the option of setting up or editing a personal vocabulary in the form of name lists, function lists, etc." (Col. 5, lines 11-15). In a specific example "call uncle Willi," "uncle Willi" is the non-command voice data segment, which is part of the speaker-dependent vocabulary (Col. 5, lines 36-41).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of storing data segments and using the data segments in executing the at least one command as taught by Stammler for Walker's method because Stammler provides the speaker dependent recognizer so that the user/speaker has the option of setting up or editing personal vocabulary in the form of name lists, function lists, etc., and adapt this vocabulary at any time to his/her needs (Col. 5, lines 13-18). This name lists and function lists (data) are necessary for executing complete commands.

As per claim 20, Walker teaches the method as recited in claim 16, but does not specifically mention wherein the step of extracting includes calling a vocabulary for recognizing numbers and recognizing the numbers in the utterance.

However, Stammler does teach wherein the step of extracting includes calling a vocabulary for recognizing numbers and recognizing the numbers in the utterance (Col. 4, lines 59-63).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of calling a vocabulary for recognition of numbers and recognizing the numbers in the utterance as taught by Stammler for Walker's method because commands requiring storing telephone numbers or changing channels require the recognizer to be able to recognize the numbers.

As per claim 23, Walker teaches the method as recited in claim 16, but does not specifically mention wherein the step of associating includes the step of changing a recognizer vocabulary and submitting at least one non-command voice data segment for recognition.

However, Stammler does teach wherein the step of associating includes the step of changing a recognizer vocabulary and submitting at least one non-command voice data segment for recognition (Col. 5, lines 33-41). The speaker dependent recognizer is connected without interface to a speaker independent recognizer. In a specific example, "call uncle Willi," the word "call" is part of the speaker independent vocabulary and "uncle Willi" is part of the speaker dependent vocabulary (Col. 5, lines 33-41), wherein "uncle Willi" is the non-command voice data segment.

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of changing a recognizer vocabulary and submitting at least one non-command voice data segment for recognition as taught by Stammler for Walker's method because Stammler provides a speech recognition unit consisting an independent compound- word recognizer and a speaker dependent additional speech recognizer (Col. 2, lines 47-49), wherein the independent recognizer recognizes general control command, numbers, names, letters, etc, and the speaker dependent recognizer recognizes user-specific/speaker-specific names or functions (non-command), which the user/speaker defines and trains (Col. 5, lines 11- 13).

8. Claims 21 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Walker (US Patent 6,434,529) in view of Kanevsky et al. (US Patent 6,434,520), hereinafter Kanevsky.

As per claim 21, Walker teaches the method as recited in claim 16, but does not specifically mention wherein the step of extracting includes extracting acoustic data based on acoustic word boundaries and saving the acoustic data for acoustically rendering the acoustic data.

However, Kanevsky does teach wherein the step of extracting includes extracting acoustic data based on acoustic word boundaries and saving the acoustic data for acoustically rendering the acoustic data (Fig. 1 and Col. 7, lines 22-30 and Col. 2, lines 1-4). An audio indexing system and method that includes a speech recognition/transcription module 109 (from Fig. 1), which stores the segmented audio data stream S1-SN 104 with the corresponding speaker identity tags IDI-ID2 106, the environment/channel tags E1-EN 108, and the corresponding transcription T1 -TN 110. Each segment may also be stored with its corresponding acoustic waveform, a subset of a few seconds of acoustic features, and/or a voiceprint, depending on the application and available memory (Col. 7, lines 22-30). Also the user may retrieve stored audio segments from the database by formulating queries based on one or more parameters corresponding to such indexed information (Col. 2, lines 1-4).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of extracting acoustic data based on acoustic word boundaries and saving the acoustic data for acoustically rendering as

taught by Kanevsky for Walker's method because Kanevsky provides an audio processing system and method for indexing and storing audio data, and an information retrieval system which provides immediate access to audio data stored in the archive through a description of the content of an audio recording, the identity of speakers in the audio recording, and/or a specification of circumstances surrounding the acquisition of the recordings (Col. 1, lines 32-38).

As per claim 22, Walker teaches the method as recited in claim 16, but does not specifically mention wherein the step of extracting includes extracting acoustic data based on acoustic word boundaries and decoding the acoustic data for storage (Fig. 1, Col. 6, lines 39-42, and Col. 7, lines 22-30). An audio indexing system and method that includes a speech recognition/transcription module 109 (from Fig. 1), which decodes the spoken utterances for each segment S1-SN 104 and generates a corresponding transcription T1-TN 110 (Col. 6, lines 39-42). The system also stores the segmented audio data stream S1-SN 104 with the corresponding speaker identity tags ID~-ID2 106 the environment/channel tags E1-EN 108, and the corresponding transcription T1-TN 110. Each segment may also be stored with its corresponding acoustic waveform, a subset of a few seconds of acoustic features, and/or a voiceprint, depending on the application and available memory (Col. 7, lines 22-30).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of extracting acoustic data based on acoustic word boundaries and decoding the acoustic data for storage as taught by

Kanevsky for Walker's method because Kanevsky provides an audio processing system and method for indexing and storing audio data, and an information retrieval system which provides immediate access to audio data stored in the archive through a description of the content of an audio recording, the identity of speakers in the audio recording, and/or a specification of circumstances surrounding the acquisition of the recordings (Col. 1, lines 32-38).

9. Claims 32-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over Walker (US Patent 6,434,529), in view of Romero (US 2002/0111803), and further in view of Stammler (US Patent 6,839,670).

As per claim 32, Walker teaches a system for recognizing commands and voice data in a same utterance comprising:

an acoustic input, which receives utterances (Fig. 1, audio input 14); and

a data buffer configured to store audio data representing the utterances (Col. 14, lines 62-67, "In the SUSPENDED state 136 (from Fig. 8) the Recognizer buffers incoming audio. This buffering allows a user to continue speaking without speech data being lost. Once the Recognizer returns to the LISTENING state the buffered audio is processed to give the user the perception of real-time processing."); and

at least one program that executes label-identified commands (Col. 13, lines 8-24).

However, Walker does not specifically mention

a speech recognition engine configured to match portions of the utterances to acoustic models and language models to recognize words and word boundaries in the utterance and labels commands in the utterance.

Conversely, Romero does teach

a speech recognition engine configured to match portions of the utterances to acoustic models and language models to recognize words and word boundaries in the utterance and labels commands in the utterance (Fig. 1, Paragraphs [0028] and [0020,0021,0022]. Speech recognizer 100 comprising an acoustic model 104 and a language model 116 (From Fig. 1). The recognizer also has a "fast acoustic match" 108, which makes use of the acoustic models (from Fig. 1), for comparing a string of incoming labels to the items stored in the conceptual vocabulary (Paragraph [0028]). Also Romero's paragraphs [0020], [0021], and [0022] show examples of “tags" (or labeling) of an utterance, such as in paragraph [0020], for the utterance "Please, give-me the phone number of Pedro Romero," the recognizer analyzes the fragment "Give me the phone number of' as a semantic identifier (command) and tagged "QUERY" or "QUERY-EN" and "Pedro Romero" as data and tagged "Pedro_fn Romero_ln."

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of a speech recognizer as taught by Romero for Walker's system because Romero provides a speech recognizer that can accept Natural Language utterances as input and directly generate the information required to process a user request (Paragraph [0007])).

However, neither Walker nor Romero specifically mention

processing remaining portions of the utterance including processing audio data parts separately from the commands using a different vocabulary, the vocabulary being selected in accordance with at least one command in the utterance.

Conversely, Stammler does teach

processing remaining portions of the utterance including processing audio data parts separately from the commands using a different vocabulary, the vocabulary being selected in accordance with at least one command in the utterance (Col. 9, lines 43-51 and Col. 5, lines 36-41 clearly provide examples where a command is processed with a speaker-independent vocabulary and accordingly the audio data is processed by a speaker-dependent vocabulary.).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of processing remaining portions of the utterance including processing audio data parts separately from the commands using a different vocabulary, the vocabulary being selected in accordance with at least one command in the utterance as taught by Stammler for Walker's system, as modified above, because Stammler provides the speaker dependent recognizer so that the user/speaker has the option of setting up or editing personal vocabulary in the form of name lists, function lists, etc., and adapt this vocabulary at any time to his/her needs (Col. 5, lines 13-18). This name lists and function lists (data) are necessary for executing complete commands.

As per claim 33, Walker, as modified above, teaches the system as recited in claim 32, wherein the at least one program includes a function which searches the utterance for labels output from the speech recognition engine to execute a command associated with the label (Walker's Col. 4, lines 41-49, " Processing of recognition results in the application program may be simplified to an invocation of the tags parser (tags parser program 24) such as

"public void interpretResult(RecognitionResult recognitionResult { TagsParser.parseResult(recognitionResult);}").

As per claim 34, Walker, as modified above, teaches the system as recited in claim 32, wherein, in accordance with each label, an audio segment is identified and processed (Walker's Col. 4, lines 43-49 describe an example of the application program processing a recognition result, wherein the recognition result could be Romero's example from Paragraph [0020]) of the tag "QUERY" representing the semantic identifier "Give me the phone number of" and the tag "Pedro_fn Romero_ln" representing the data of the utterance "Please, give me the phone number of Pedro Romero.").

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used examples of Natural Language utterances as taught by Romero for Walker's system because Romero provides a speech recognizer that can

accept Natural Language utterances as input and directly generate the information required to process a user request (Paragraph [0007])).

As per claim 35, Walker, as modified above, teaches the system as recited in claim 32, wherein the speech recognition engine utilizes grammars with labels, which the system uses for assigning labels to decoded commands (Walker's Col. 4, lines 34-40, show an example of the rule grammar applied to a media-player application, wherein, for example, the system assigns the label to the decoded commands (play|go|start)).

As per claim 36, Walker, as modified above, teaches the system as recited in claim 35, wherein the grammars are represented in Bachus-Naur Form (BNF) (Walker's Fig. 4).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to NATALIE LENNOX whose telephone number is (571)270-1649. The examiner can normally be reached on Monday to Friday 9:30 am - 7 pm (EST).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)272-7602. The fax phone

number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

NL 08/20/2008

/Richemond Dorvil/
Supervisory Patent Examiner, Art Unit 2626